## **CLAIMS**

Having thus described my invention, what I claim as new and desire to secure by Letters Patent is as follows:

1	1. A method of classifying a source publishing a document on a portion
2	of a network, comprising steps of:
3	electronically receiving a document;
4	based on the document, determining a source which published the
5	document; and
6	assigning a code to said document based on whether data
<b>7</b>	associated with the document published by the source matches with data
8	contained in a database.
1	2. The method according to claim 1, wherein said portion of said
2	network comprises a graphical multimedia portion of said network, said
3	source comprises a Web site publishing a home page, and said network
4	comprises the Internet.
1	3. The method according to claim 2, wherein said graphical multimedia
2 ·	portion of said network comprises the World-Wide Web (WWW) and
3	said document comprises a Web document,
4	wherein said step of assigning a code includes determining that
5	the Web site comprises a first entity when there is a match of the Web
5	site with said data, and determining that the Web site comprises a second
7	entity when there is no match of the Web site with said data.
l	4. The method according to claim 1, wherein said step of determining a
2	source includes:
3	extracting a domain name from a predetermined uniform
ļ	resources locator (URL) database;

1	querying a database for storing registered domain names; and
2	merging an address database with predetermined data.
	·
1	5. The method according to claim 4, wherein said predetermined data
2	comprises Yellow Pages data,
3	wherein said step of determining further comprises:
4	characterizing uniform resource locators (URLs) by their
5	Internet Protocol (IP) addresses including identifying a plurality of
6	attributes based on the IP addresses of new URLs, a new URL being
7	retrieved and parsed into a domain name and directory path portions, and
8	determining, based on said domain name, whether a
9	selected URL is hosted on one of a true server and a shared server.
1	6. The method according to claim 5, said step of determining further
2	comprising:
3	for a shared server, determining a root path by searching for the
4	given domain name in a new URL database and identifying common
5	directory paths,
6	wherein, when no match is present, the URL is processed
7	subsequently at a later iteration, and, when a match is present, the root
8	path is set to a matching path.
1	7. The method according to claim 6, wherein said step of assigning a
2	code comprises:
3	automatically identifying a business associated with the source
4	publishing said document, said business being hosted on a Service
5	Provider (SP) Web server.
1	8. The method according to claim 7, wherein said step of assigning a
2	code further comprises:
3	receiving a URL based on said determining step; and

4	a URL determining step for determining whether said URL
5	comprises one of a root URL and a leaf URL.
1	O The method according to the control of the contro
2	9. The method according to claim 8, wherein said root URL comprises
3	an entry point for a home page on the World-Wide Web, and a leaf UR
4	comprises a link below a root URL,
5	wherein said URL determining step comprises:
	parsing said URL into a domain name component and a
6	directory path component;
7	analyzing the domain name in said domain name
8	component to determine whether it is associated with an SP;
9	when the domain name is not associated with an SP,
10	checking the directory path component to judge whether a directory path
1	is missing, a missing directory path indicating a root URL;
2	when the domain name is associated with an SP, checking
.3	whether a directory path does not exist to thereby determine that said
4	domain name comprises a root URL, and when a directory path exists,
5	then comparing the path to known SP Client Directory paths.
1	10. The method according to claim 9, further comprising:
2	when said URL is determined to be a root URL, analyzing a
3	home page associated with said root URL automatically to extract home
4	page data contained therein and assigning the home page data to the Roo
5	URL being analyzed.
1	11. The method according to claim 10, further comprising:
2	comparing said home page data with data in a predetermined
3	business organizations database,
1	wherein, when there is a match, said code is assigned to the
5	corresponding root URL, and, when no match is found, said URL is
5	identified for subsequent analysis of lower-level hyperlinks during a paye
-	

7.	iteration of said method.
1	12. The method according to claim 11, wherein when no match is found
2	at any level, said home page is identified as a personal page.
1	13. A method of automatically assigning a document a code for
2	distinguishing a first-type page from a second-type page, comprising
3	steps of:
4	electronically receiving a document;
5	based on the document, determining a source which published the
6	document; and
7	assigning a code to said document based on whether the source
8 .	matches with data contained in a database.
1	14. A search engine for use on a network for distinguishing between
2	business web pages and personal web pages, comprising:
3	means for parsing the content of a hyper-text markup language
4	(HTML) at a web address and searching for criteria contained therein;
5	means for analyzing a uniform resources locator (URL) of the
6	web address to determine characteristics thereof of a web page at the web
7	address;
8	means for determining whether said criteria match with data
9	contained in a database; and
10	means for cross-referencing a match, determined by said
11	determining means, to a second database, to classify a source which
12	published the web page.
1	15. A search engine according to claim 14, wherein said criteria include
2	at least one of an address, a telephone numbers, a facsimile number, a
	,

contact and a key-word contained in said HTML, and

3

4	wherein the characteristics of said web page include a
5	geographical location and a web host computer.
	* **
1	16. A search engine according to claim 14, wherein said database
2	includes a Business Semantic Terminology database having information
3	related to business categories in a Yellow Pages directory.
1	17. A search engine according to claim 14, wherein said second database
2	includes a Yellow Pages database.
1	18. A search engine according to claim 14, wherein said web page
2	comprises hyperlinks, and said means for parsing comprises an indexer
3	robot for traversing said hyperlinks in said web page and a web index
4	database,
5	said indexer robot for indexing a content of said web page into
6	said web index database.
1	19. A search engine according to claim 14, wherein said means for
2	analyzing comprises:
3	means for determining whether said URL comprises one of a root
4	URL and a leaf URL.
1 .	20. A search engine according to claim 19, wherein said root URL
2	comprises an entry point for the web page on the World-Wide Web, and
3	a leaf URL comprises a link below a root URL, said search engine
4	further comprising:
5	means for parsing said URL into a domain name component and a
5	directory path component;
7	means for analyzing the domain name in said domain name
3	component to determine whether it is associated with an SP;

9	means for checking the directory path component to judge
10	whether a directory path is missing, when the domain name is not
11	associated with a service provider (SP), a missing directory path
12	indicating a root URL, and for checking whether a directory path does
13	not exist to thereby determine that said domain name comprises a root
14	URL, when the domain name is associated with an SP;
15	means for comparing the path to known SP Client Directory
16	paths, when a directory path exists;
17 ·	means for analyzing a home page associated with said root URL,
18	when said URL is determined to be a root URL, thereby automatically to
19	extract home page data contained therein; and
20	means for assigning the home page data to the Root URL being
21	analyzed.